

Trempler, Kati; Hetmanek, Andreas; Wecker, Christof; Kiesewetter, Jan; Wermelt, Mia; Fischer, Frank; Fischer, Martin; Gräsel, Cornelia

Nutzung von Evidenz im Bildungsbereich. Validierung eines Instruments zur Erfassung von Kompetenzen der Informationsauswahl und Bewertung von Studien

Blömeke, Sigrid [Hrsg.]; Zlatkin-Troitschanskaia, Olga [Hrsg.]: *Kompetenzen von Studierenden*. Weinheim u.a. : Beltz Juventa 2015, S. 144-166. - (Zeitschrift für Pädagogik, Beiheft; 61)



Quellenangabe/ Reference:

Trempler, Kati; Hetmanek, Andreas; Wecker, Christof; Kiesewetter, Jan; Wermelt, Mia; Fischer, Frank; Fischer, Martin; Gräsel, Cornelia: *Nutzung von Evidenz im Bildungsbereich. Validierung eines Instruments zur Erfassung von Kompetenzen der Informationsauswahl und Bewertung von Studien* - In: Blömeke, Sigrid [Hrsg.]; Zlatkin-Troitschanskaia, Olga [Hrsg.]: *Kompetenzen von Studierenden*. Weinheim u.a. : Beltz Juventa 2015, S. 144-166 - URN: urn:nbn:de:0111-pedocs-155089 - DOI: 10.25656/01:15508

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-155089>

<https://doi.org/10.25656/01:15508>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

61. Beiheft

April 2015

ZEITSCHRIFT FÜR PÄDAGOGIK

**Kompetenzen
von Studierenden**

BELTZ JUVENTA

Zeitschrift für Pädagogik · 61. Beiheft

Kompetenzen von Studierenden

Herausgegeben von

Sigrid Blömeke und Olga Zlatkin-Troitschanskaia

BELTZ JUVENTA

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, bleiben dem Beltz-Verlag vorbehalten.

Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, bei der die einzelnen Zahlungsmodalitäten zu erfragen sind.

© 2015 Beltz Juventa · Weinheim und Basel

www.beltz.de · www.juventa.de

Herstellung: Lore Amann

Satz: text plus form, Dresden

E-Book

ISSN 0514-2717

Bestell-Nr. 443508

Inhaltsverzeichnis

<i>Sigrid Blömeke/Olga Zlatkin-Troitschanskaia</i> Kompetenzen von Studierenden. Einleitung zum Beiheft	7
--	---

<i>Lars Jenßen/Simone Dunekacke/Sigrid Blömeke</i> Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis	11
---	----

Berufsbezogene Kompetenzen

<i>Svenja Hammer/Sonja A. Carlson/Timo Ehmke/Barbara Koch-Priewe/ Anne Köker/Udo Ohm/Sonja Rosenbrock/Nina Schulze</i> Kompetenz von Lehramtsstudierenden in Deutsch als Zweitsprache: Validierung des GSL-Testinstruments	32
--	----

<i>Josef Riese/Christoph Kulgemeyer/Simon Zander/Andreas Borowski/ Hans E. Fischer/Yvonne Gramzow/Peter Reinhold/Horst Schecker/ Elisabeth Tomczyszyn</i> Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik	55
--	----

<i>Simone Dunekacke/Lars Jenßen/Sigrid Blömeke</i> Mathematikdidaktische Kompetenz von Erzieherinnen und Erziehern: Validierung des KomMa-Leistungstests durch die videogestützte Erhebung von Performanz	80
--	----

<i>Franziska Bouley/Stefanie Berger/Sabine Fritsch/Eveline Wuttke/ Jürgen Seifried/Kathleen Schnick-Vollmer/Bernhard Schmitz</i> Der Einfluss von universitären und außeruniversitären Lerngelegenheiten auf das Fachwissen und fachdidaktische Wissen von angehenden Lehrkräften an kaufmännisch-berufsbildenden Schulen	100
--	-----

<i>Olga Zlatkin-Troitschanskaia/Manuel Förster/Susanne Schmidt/ Sebastian Brückner/Klaus Beck</i> Erwerb wirtschaftswissenschaftlicher Fachkompetenz im Studium – Eine mehrbenenanalytische Betrachtung von hochschulischen und individuellen Einflussfaktoren	116
---	-----

Gabriele Kaiser

Erfassung berufsbezogener Kompetenzen von Studierenden.

Ein Kommentar 136

Forschungsbezogene Kompetenzen

Kati Trempler/Andreas Hetmanek mit Christof Wecker/Jan Kiesewetter/

Mia Wermelt/Frank Fischer/Martin Fischer/Cornelia Gräsel

Nutzung von Evidenz im Bildungsbereich – Validierung

eines Instruments zur Erfassung von Kompetenzen

der Informationsauswahl und Bewertung von Studien 144

Sandra Schladitz/Jana Groß Ophoff/Markus Wirtz

Konstruktvalidierung eines Tests zur Messung

bildungswissenschaftlicher Forschungskompetenz 167

Alexandra Winter-Hözl/Kristin Wäschle/Jörg Wittwer/

Rainer Watermann/Matthias Nückles

Entwicklung und Validierung eines Tests zur Erfassung

des Genrewissens Studierender und Promovierender

der Bildungswissenschaften 185

Gabriele Steuer/Tobias Engelschalk/Gregor Jöstl/Anne Roth/

Bastian Wimmer/Bernhard Schmitz/Barbara Schober/Christiane Spiel/

Albert Ziegler/Markus Dresel

Kompetenzen zum selbstregulierten Lernen im Studium:

Ergebnisse der Befragung von Expert(inn)en aus vier Studienbereichen 203

Johannes König

Stand der Forschung zu wissenschaftsbezogenen Kompetenzen

und weiterführende Fragen. Ein Kommentar 226

Forschungsbezogene Kompetenzen

*Kati Trempler/Andreas Hetmanek mit Christof Wecker/Jan Kiesewetter/
Mia Wermelt/Frank Fischer/Martin Fischer/Cornelia Gräsel*

Nutzung von Evidenz im Bildungsbereich

*Validierung eines Instruments zur Erfassung von Kompetenzen
der Informationsauswahl und Bewertung von Studien*

Zusammenfassung: Evidenzbasierte Praxis kann nach dem Vorbild der Medizin als reflektierte Nutzung der besten verfügbaren empirischen Befunde für die berufliche Tätigkeit beschrieben werden. Dies stellt eine komplexe Anforderung für pädagogisches Personal dar. Ziel der vorliegenden Arbeit ist die Validierung eines Instruments zur Erfassung der dafür benötigten Kompetenz. Im Fokus stehen die beiden Teilkompetenzen *Informationsauswahl* und *Bewertung von Studien*. An der Studie nahmen 341 Studierende aus erziehungswissenschaftlichen Studiengängen teil. Es wurde ein fallbasierter Online-Test für Informationsauswahl und Bewertung von Studien durchgeführt. Die Ergebnisse liefern Hinweise für die Trennbarkeit der beiden Teilkompetenzen und zeigen plausible Zusammenhänge zu weiteren Variablen.

Schlagworte: Validierung, Kompetenzmessung, fallbasierte Entscheidung, evidenzbasierte Praxis im Bildungsbereich, Primärstudien

1. Einleitung

Die alltäglichen Anforderungssituationen, mit denen Praktiker im Bildungsbereich konfrontiert werden, sind vielfältig und umfassen die Organisation und Aufrechterhaltung von Aktivitäten innerhalb sozialer Gruppen, die Entwicklung und Verbreitung von Inhalten sowie die zeitliche Planung von Abläufen (Bromme, 2014, S. 89). Zahlreiche leicht zugängliche ratgebende und anleitende Materialien zur Gestaltung dieser Situationen existieren; vor allem in Form von Zeitschriften, Handbüchern und im Internet. Qualitativ hochwertige und wissenschaftlich abgesicherte Literatur allerdings ist selten darunter zu finden. Dabei kann sie durchaus entscheidend zur Bewältigung von Anforderungen der beruflichen Praxis beitragen (Bromme, Prenzel, & Jäger, 2014, S. 8). Relevante wissenschaftliche Forschungsarbeiten zu finden, auszuwählen sowie mit Blick auf eine konkret anstehende Entscheidung zu bewerten und nutzbar zu machen, stellt eine komplexe Anforderung für pädagogisches Personal dar. Eine reflektierte und auf-

geklärte evidenzbasierte Praxis im Bildungsbereich kann sich aber nur dann entwickeln, wenn diese Anforderung von verantwortlichen Akteuren bewältigt wird. Die *Informationsauswahl* und die *Bewertung von Studien* sind zwei wichtige Bestandteile der komplexen Kompetenz des evidenzbasierten Argumentierens. Im Mittelpunkt dieses Beitrages stehen die theoretische Modellierung und die Validierung eines Tests für die beiden Teilkompetenzen.

2. Hintergrund

2.1 Evidenzbasierung im Bildungsbereich

Ihren Ursprung hat die Evidenzbasierung im Bereich der Medizin. Als neues Paradigma der medizinischen Ausbildung und Patientenversorgung wurde bereits vor mehr als zwei Jahrzehnten durch die Evidence-Based Medicine Working Group (1992) die evidenzbasierte Praxis eingeführt. Die Idee, versorgungsrelevante Fragen stärker auf Grundlage vorhandener wissenschaftlicher Erkenntnisse in Form von wissenschaftlicher Primärliteratur und weniger auf Grundlage individueller bzw. tradierter Erfahrung zu bearbeiten, wird zur Erfolgsgeschichte (u. a. Montori & Guyatt, 2008, S. 1814). Erste Versuche, Unterricht evidenzbasiert zu gestalten, finden sich konsequenterweise in der medizinischen Ausbildungsforschung (Harden, Grant, Buckley & Hart, 1999).

Seit der Veröffentlichung der ersten PISA-Ergebnisse 2000 (Baumert et al., 2002) wird ein ähnlicher Paradigmenwechsel auch für den Bildungsbereich diskutiert (u. a. Bromme et al., 2014; Slavin, 2008) und teilweise aktiv politisch vorangetrieben (z. B. KMK, 2014, S. 3). Im Mittelpunkt der Diskussion steht die Frage, in welcher Weise bildungswissenschaftliche Befunde für die Praxis gesucht, beurteilt und verwendet werden können – sowohl auf der Ebene politischer Steuerung wie auch auf der Ebene von Entscheidungen einzelner Personen (Bromme et al., 2014, S. 6). Damit die Forderung nach evidenzbasierter Praxis im Bildungsbereich nach dem Vorbild der Medizin fruchtbar werden kann, muss ein kompetenter Umgang mit originär wissenschaftlicher Literatur als Grundvoraussetzung für eine evidenzbasierte Entscheidungskultur durch die im Bildungsbereich tätigen Praktiker (z. B. Erzieherinnen und Erzieher, Lehrkräfte an allen Schulformen, Fort- und Weiterbildner) gegeben sein. Hierfür muss der Umgang mit wissenschaftlicher Literatur in das professionelle Berufsverständnis integriert werden (Tugwell, Haynes & Sackett, 1992).

Für den Bildungsbereich wird evidenzbasierte Praxis nach dem Vorbild der Medizin als reflektierte Nutzung der besten verfügbaren und empirisch gesicherten Evidenz für das Treffen und die Begründung von beruflichen Entscheidungen definiert (vgl. Bromme et al., 2014, S. 9; Sackett, Rosenberg, Gray, Haynes & Richardson, 1996, S. 71). Damit stellt sie einen Gegenentwurf zur verbreiteten Praxis dar, Entscheidungen überwiegend auf Basis von individuellen Erfahrungen, Ideologien oder subjektiven Überzeugungen zu treffen (Harden et al., 1999; Slavin, 2008). Diskussionen zur Evidenzbasierung im Bildungsbereich befassen sich bislang vor allem mit der Ebene der evidenzbasierten

Steuerung, also der Nutzung wissenschaftlicher Evidenz als Grundlage für bildungspolitische Entscheidungen (*evidence-based reform*; Slavin, 2008). Eine weitere Ebene der Nutzung von Evidenz als Grundlage für individuelles professionelles Handeln im Bildungsbereich findet bislang wenig Beachtung (Bromme et al., 2014), obwohl beispielsweise bereits im Jahr 2004 durch die Kultusministerkonferenz die Standards für die Lehrerbildung vorgelegt wurden, in denen die Planung, Organisation und Reflexion von Lehr-Lern-Prozessen anhand bildungswissenschaftlicher Erkenntnisse als „Kernaufgabe“ von Lehrkräften definiert wird (KMK, 2014, S. 3). Diese Kernaufgabe erfordert Fähigkeiten der Rezeption und Bewertung bildungswissenschaftlicher Forschungsergebnisse sowie der Nutzbarmachung von Ergebnissen der Bildungsforschung für die Tätigkeit als Lehrkraft (KMK, 2014, S. 13). Bildungswissenschaftliche Evidenz umfasst empirisch abgesicherte Erkenntnisse, die aus qualitativ hochwertigen Untersuchungen der Bildungsforschung gewonnen werden; sie ermöglicht es, Hinweise für die Wirksamkeit pädagogischer Maßnahmen zu extrahieren und pädagogische Entscheidungen auf der Grundlage von abgesichertem Wissen zu rechtfertigen (Cook, Smith & Tankersley, 2012, S. 498 ff.). Welche Art von Forschung hierfür besonders geeignet ist, bildet Gegenstand fortlaufender Diskussion (Beelmann, 2014; Cook & Gorard, 2007; Fischer, Waibel & Wecker, 2005; Wecker, 2013). Auch wenn systematische Forschungssynthesen in der Regel höheren Qualitätsansprüchen genügen (siehe Diskussion bei Beelmann, 2014), sind empirische Forschungsartikel die primäre Quelle für forschungsbasierte Erkenntnisse: Zum einen gibt es zu vielen relevanten Themenkomplexen im Bildungsbereich bislang kaum systematische Reviews oder Metaanalysen, zum anderen gehen in derartigen Überblicksarbeiten notwendigerweise viele umsetzungsrelevante Details verloren. Damit ist die Fähigkeit, kritisch und kompetent mit empirischer Primärliteratur umzugehen, als von grundsätzlicher Bedeutung für evidenzbasierte Praxis anzusehen (EBMWG, 1992). Aus diesem Grund befasst sich die vorliegende Untersuchung mit dem Umgang von Personen im Bildungsbereich mit empirischen Studien.

2.2 Die Kompetenz im evidenzbasierten Argumentieren

Im Folgenden werden zunächst anhand einer alltäglichen Anforderungssituation im Bildungsbereich zwei mögliche Handlungsalternativen vorgestellt, die den Paradigmenwechsel von aktuell gängigen Verfahren der professionellen pädagogischen Praxis zu einer stärker evidenzbasierten Praxis im Bildungsbereich skizzieren. Daran anschließend wird das im Rahmen des hier vorgestellten Forschungsvorhabens entwickelte Prozessmodell für Leistungen beim evidenzbasierten Argumentieren vorgestellt. Genauere Spezifikationen des Zusammenspiels der beiden untersuchten Teilkompetenzen mit weiteren Komponenten bzw. Faktoren und das Vorgehen bei der Validierung des Testinstrumentes zur Erfassung der Leistungen werden im Anschluss daran präsentiert.

Die Anforderungssituation

Alltagsszenario: zu einer der häufigsten Aufgaben von pädagogischen Praktikern im Bildungsbereich gehört die Planung und Anleitung von Aktivitäten zum Erwerb von Wissen und Fertigkeiten. Die Frage ist nun, wie der Einzelne dieser Anforderung begegnet. Wie geht er bei der Entwicklung eines effektiven Konzepts für Wissensvermittlung bzw. Fertigkeitserwerb vor?

Der aktuell gängige Weg: Der pädagogische Praktiker denkt über den Inhalt bzw. das Lehrziel nach und überlegt, wie er es am besten erreichen könnte. Zusätzlich sichtet er alte Konzepte mit ähnlichen Inhalten bzw. Zielen, spricht eventuell mit weiteren Kolleginnen, konsultiert einen „methodischen Werkzeugkasten“ im Internet oder in Buchform und blättert – sofern genug Zeit zur Verfügung steht – in den Unterlagen aus dem eigenen Studium oder Materialien aus der Ausbildung. Dabei entsteht ein Vermittlungskonzept, das überwiegend inhaltsfokussiert ist und methodisch auf unsystematischen Beobachtungen und Erfahrungen sowie Intuition aufbaut.

Der Weg der evidenzbasierten Praxis: Der verantwortliche Pädagoge analysiert den zu vermittelnden Lehrinhalt, die Zielgruppe sowie weitere wichtige Einflussgrößen. Insofern ihm noch kein evidenzbasiertes Wissen zu diesem Typ von Anforderungssituation zur Verfügung steht, überlegt er sich Suchbegriffe, unter denen wissenschaftliche Erkenntnisse zu möglichst ähnlichen Typen von Lerninhalten bei ähnlicher Ausgangslage publiziert sein könnten, und sucht systematisch in Fachportalen nach passenden empirischen Untersuchungen, wählt die wichtigsten (relevantesten und qualitativ besten) aus und liest und bewertet sie systematisch anhand eines elaborierten Bewertungsschemas. Dabei achtet er bspw. auf die Passung der Studie zu seinen Lehrzielen und den gegebenen Bedingungen sowie auf die Qualität der Studien. Auf Grundlage dieser Informationen entwickelt er sein Vermittlungskonzept. Bei der Dokumentation zieht er die gewählte Primärliteratur als Begründung für das gewählte Vorgehen heran. Bei den zu treffenden Entscheidungen kommt systematisch abgesichertem Wissen überwiegendes Gewicht zu.

So könnte sich der Paradigmenwechsel im Bildungsbereich nach dem Vorbild der Medizin vollziehen. In der hier bewusst überzeichneten „Reinform“ ist und bleibt diese Art der Evidenzbasierung Fiktion – in der Medizin gleichermaßen wie im Bildungsbereich. In beiden Bereichen werden aktuell lebhaft Debatten um eine sinnvolle, praktikierbare und zielführende Ausgestaltung geführt (Pant, 2014, S. 81).

In jedem Fall wird für die Umsetzung einer stärker evidenzbasierten Praxis vom professionellen Praktiker ein ganzes Bündel von neuen Fertigkeiten verlangt. Besonders grundlegend sind die effiziente Literatursuche und -auswahl sowie die Anwendung von Regeln zur Bewertung von empirischer Forschung (EBMWG, 1992, S. 2420). Weiterhin müssen die abgesicherten Erkenntnisse aus der Forschung mit anderen Faktoren (Praktikabilität, Erfahrungen mit der spezifischen Gruppe von Lernenden etc.) integriert und bei der Entscheidung berücksichtigt sowie reflektiert, dokumentiert und oftmals auch kommuniziert werden. Dieser Ablauf stellt eine komplexe und vielschichtige kognitive Anforderung dar, für deren Bewältigung verschiedene Teilfertigkeiten, spezifisches Wissen und Einstellungen zusammenspielen und verschiedene Phasen kognitiver

Aktivität durchlaufen werden müssen. Als Bezeichnung für diesen Problemlöseprozess führen wir den Begriff „evidenzbasiertes Argumentieren“ ein. Beim evidenzbasierten Argumentieren und den spezifischen Anforderungen der *Informationsauswahl* und *Bewertung von Studien* handelt es sich um gering strukturierte Probleme (Simon, 1973). Die Entwicklung eines theoretischen Modells für den beschriebenen Prozess ist eine zentrale Leistung des zugrunde liegenden Forschungsvorhabens.

Zur theoretischen Modellierung der Kompetenz im evidenzbasierten Argumentieren im Bildungsbereich greifen wir Konzeptionen aus der Medizin (Rosenberg & Donald, 1995, S. 1123) auf und erweitern diese mit Theorien und Erkenntnissen aus der Forschung zum *Information Problem Solving* (im Folgenden „IPS“; Brand-Gruwel, Wopereis & Vermetten, 2005) für den Bereich der systematischen Informationsbeschaffung, -bewertung und -integration und zum *Evidence-based Reasoning Framework* (u. a. Brown, Furtak, Timms, Nagashima & Wilson, 2010) für den Teilbereich der Entscheidungsfindung und argumentativen Begründung.

Aufgrund der besonders grundlegenden Bedeutung konzentrieren wir uns in diesem Beitrag auf die beiden Teilbereiche (1) *Informationsauswahl* und (2) *Bewertung von Studien*.

Informationsauswahl

Der erste Schritt zur Bewältigung der Anforderungen der evidenzbasierten Praxis ist das Suchen und Auswählen aussagekräftiger Untersuchungen, um ein Informationsproblem zu lösen. Dieser Schritt wird üblicherweise anhand von Datenbanken oder in Internetsuchmaschinen durchgeführt. Von Brand-Gruwel, Wopereis und Walraven (2009) wurde ein ausführliches Modell zur Lösung von Informationsproblemen vorgelegt und empirisch bestätigt. Zentraler Bestandteil dieses Modells sind die iterativen Prozesse von Informationssuche, Überfliegen und tiefer Verarbeitung der Information (S. 1208). Für das Gelingen dieser Prozesse ordnen die Autoren den regulierenden Aktivitäten Orientierung, Monitoring, Steuerung und Prozessevaluation eine zentrale Rolle zu (S. 1209). Experten-Novizen-Vergleiche zeigen, dass Experten deutlich mehr Zeit mit vertiefter Beschäftigung mit dem Inhalt verbringen als Novizen (Brand-Gruwel et al., 2005, S. 498). Dies deckt sich mit Erkenntnissen aus einem weiteren relevanten Forschungszweig zum Verständnis multipler Quellen (*Multiple Source Comprehension*; u. a. Wiley et al., 2009). Auch hier spielen das Verstehen und die Bewertung der Relevanz und Nützlichkeit einer Information eine zentrale Rolle (Goldman, Lawless, Pellegrino, Braasch & Gomez, 2012, S. 182). Diese Reflexion und Bewertung sind insbesondere notwendig bei einer gegebenen Fülle an Informationen, die beispielsweise bei einer Internetsuche in Form von Listen ungefiltert – d. h. ohne jegliche Form der Plausibilitäts- und Qualitätsprüfung – ausgegeben werden. Für derartige Tätigkeiten sind spezifische Wissensbestände zur Bewertung der Relevanz und Qualität der Informationen notwendig, damit erfolgreich die aussagekräftigsten Forschungserkenntnisse aufgefunden werden können. Die Wichtigkeit von anforderungsspezifischem Vorwissen sowie epistemologischen Überzeugungen wird auch an anderer Stelle betont (Lazonder & Rouet, 2008) und findet in einem systematischen Re-

view der Forschungsarbeiten zum IPS Unterstützung (Wopereis & van Merriënboer, 2011, S. 234–235).

Ausgehend von diesen Vorarbeiten modellieren wir die Informationsauswahl im evidenzbasierten Argumentieren als Prozess von Suche, Überfliegen und Berücksichtigung weiterer Informationen, bei dem darüber hinaus anforderungsspezifisches Vorwissen sowie epistemologische Überzeugungen eine bedeutsame Rolle spielen. Als anforderungsspezifische Wissensbestände nehmen wir Grundkenntnisse (Terminologie und Zusammenhänge) aus der pädagogischen und psychologischen Lehr-Lern-Forschung sowie zur empirischen Forschungsmethodik an.

Bewertung von Studien

Nachdem wissenschaftliche Forschungsarbeiten auf Grundlage von schnell überblickbaren Informationen (bspw. Schlagwörter und Zusammenfassungen in einer Literaturliteraturdatenbank) als relevant eingestuft und ausgewählt wurden, müssen die Publikationen einer systematischen Beurteilung unterzogen werden. Für den medizinischen Bereich haben Harden et al. (1999) ein Bewertungsraster entwickelt, mit dem sechs Aspekte wissenschaftlicher Untersuchungen abgedeckt werden: Qualität, Übertrag- und Anwendbarkeit des Behandlungsansatzes auf den aktuellen Fall, Reichweite, Stärke und Validität der Befunde sowie Vergleichbarkeit der Rahmenbedingungen. Für den Bildungsbereich fehlt bislang ein derartiges Beurteilungsschema.

Zur Entwicklung eines für den Bildungsbereich adäquaten Bewertungsschemas haben wir das Schema aus der Medizin als Ausgangspunkt herangezogen und durch Qualitätsaspekte aus der pädagogisch-psychologischen Methodenlehre erweitert und geschärft. Zu diesem Zweck haben wir zusätzlich die vier Aspekte von Validität nach Cook und Campbell (1979) sowie das UTOS-Bewertungsschema zur Übertragbarkeit von Studienergebnissen nach Cronbach (1982) integriert. Dabei entstand ein Beurteilungsschema zur systematischen Bewertung von Relevanz für eine konkrete Anwendungssituation und allgemeiner methodischer Qualität einer wissenschaftlichen Studie (siehe Abb. 3).

Der pädagogische Praktiker wird bei der (systematischen) Bewertung von Studien mit einer komplexen Anforderung konfrontiert. Zur Bewältigung muss ein komplizierter kognitiver Prozess realisiert werden, der das Lesen und Verstehen der Studie sowie letztlich das Beurteilen der einzelnen Dimensionen der wissenschaftlichen Arbeit umfasst. Es ist plausibel anzunehmen, dass umfangreiche inhaltliche und forschungsmethodische Wissensbestände im Zusammenspiel mit epistemologischen Überzeugungen und der Motivation zentral für die Bewältigung der Aufgabe sind.

Hintergrundvariablen

Grundsätzlich werden kognitive Leistungen von *allgemeinen kognitiven Fähigkeiten* (u. a. Amelang, 2000) und *spezifischem Wissen* (Feltovich, Prietula & Ericsson, 2006) beeinflusst, und Leistungsunterschiede bei komplexen kognitiven Aufgaben wie dem evidenzbasierten Argumentieren können teilweise durch Unterschiede in diesen beiden Variablen erklärt werden (Klieme, Funke, Leutner, Reimann & Wirth, 2001, S. 189;

Mayer, 2003, S. 263 ff.). Die Anforderungen bei der *Informationsauswahl* und der *Bewertung von Studien* können zudem nur bewältigt werden, wenn anforderungsspezifisches – also pädagogisch-psychologisches und forschungsmethodisches – Wissen in ausreichendem Maß vorhanden ist sowie eine ausreichende Ausprägung insbesondere verbaler Intelligenz gegeben ist.

Die Struktur der Anforderungen an den evidenzbasiert agierenden Praktiker legt es nahe, dass einige kognitive Aktivitäten Überlappungen mit denjenigen beim *wissenschaftlichen Denken* haben. Wissenschaftliches Denken umfasst die bewusste und kontrollierte Anwendung wissenschaftlicher Methoden und Prinzipien für die Koordination von Theorie und wissenschaftlichen Befunden in Argumentations- und Problemlösesituationen (Zimmerman, 2000, S. 114). Leistungsunterschiede in den Bereichen *Informationsauswahl* und *Bewertung von Studien* könnten daher möglicherweise ebenfalls durch Unterschiede in der Ausprägung dieser Fertigkeit erklärt werden.

Dies gilt auch für die kognitiven Prozesse bei der Prüfung der Zuverlässigkeit und Genauigkeit einer Aussage, der Unterscheidung zwischen relevanten und nicht relevanten Informationen sowie beim wissenschaftlich-analytischen Schlussfolgern. Daher nehmen wir an, dass Fertigkeiten des *kritischen Denkens* (Astleitner, Brünken & Zander, 2002) eine weitere relevante Einflussgröße beim Bewältigen von Anforderungen an den evidenzbasierten Praktiker darstellen.

Es wurde bereits herausgearbeitet, dass auch persönliche Überzeugungen eine Rolle spielen; so ist anzunehmen, dass *epistemologische Überzeugungen* zu Leistungen im evidenzbasierten Argumentieren beitragen. Sie werden beschrieben als „(...) intuitive Theorien, die die Art der Begegnung mit der erkennbaren Welt vorstrukturieren“ (Baumert & Kunter, 2006, S. 498). Diese Einstellungen zu Wissen und Wissenserwerb beeinflussen das Lernen, Denken und Schlussfolgern (Bromme et al., 2014; Hofer, 2000; Hofer & Pintrich, 1997). Auch die subjektive Einschätzung der Vertrautheit mit forschungsmethodischen Begriffen, also *Überzeugungen zum eigenen forschungsmethodischen Wissen*, sollte Unterschiede in Leistungen bei *Informationsauswahl* und *Bewertung von Studien* im Kontext von evidenzbasierter Praxis anteilig erklären können. Diese Überzeugungen umfassen Einschätzungen der eigenen Fähigkeiten des Umgangs mit forschungsmethodischen Inhalten und könnten sich ebenfalls in Leistungen des evidenzbasierten Argumentierens niederschlagen.

Da bislang kaum Erkenntnisse zu Einflussgrößen im Kontext gering strukturierter Probleme (Simon, 1973) beim Erbringen von Leistungen in diesem Zusammenhang vorliegen, sind Vorhersagen exakter (kausaler) Zusammenhänge und Abhängigkeiten nicht möglich.

2.3 Vorgehensweise bei der Validierung

Inhaltsvalidierung: Belege für die Inhaltsvalidität werden in aller Regel nicht quantitativ empirisch, sondern theoretisch erbracht und beziehen sich darauf, ob die Inhalte von Items das untersuchte Attribut abbilden (vgl. Hartig, Frey & Jude, 2012). Die Inhalts-

validierung umfasst in der Regel die Begutachtung der konzipierten Items eines Tests durch Experten, deren Expertise systematisch nachgewiesen werden sollte (Jenßen, Dunekacke & Blömeke, 2015, in diesem Beiheft). Wir betrachten die Inhaltsvalidität dann als gegeben, wenn die befragten Experten in ihrer Bearbeitung der Aufgaben zur Informationsauswahl sowie der Bewertung von Studien eine hohe Übereinstimmung aufweisen. Bei den Aufgaben zur *Informationsauswahl* ist die Übereinstimmung dann gegeben, wenn Experten sich über die Relevanz von Studien einig sind, d. h. dieselben Studien als relevant, weniger relevant und nicht relevant einschätzen. Bei der *Bewertung von Studien* kann von Übereinstimmung gesprochen werden, wenn Experten Studien anhand der oben genannten Qualitätsaspekte kongruent bewerten.

Konstruktvalidierung: Im Sinne von Cronbach und Meehl (1955) dient die Konstruktvalidierung dem Nachweis, welche latente Variable (Kompetenz) das Kriterium (Performanz) bestimmt. Prozeduren für die Bestimmung der Konstruktvalidität (nach Cronbach & Meehl, 1955, S. 287–288) umfassen unter anderem (1) Korrelationen zwischen den einzelnen Bereichen innerhalb der Gesamtkompetenz (Korrelation zwischen den Bereichen *Informationsauswahl* und *Bewertung von Studien*) und (2) Aussagen über die interne Struktur bzw. Konsistenz der Teilkompetenzen, welche durch Inter-Item-Korrelationen nachgewiesen werden kann (interne Konsistenz der Bereiche *Informationsauswahl* und *Bewertung von Studien*). Aufgrund der inhaltlichen Nähe (überlappende kognitive Aktivitäten) und klaren theoretischen Unterschiedenheit werden Maße der Intelligenz, spezifisches Wissen, wissenschaftliches und kritisches Denken, epistemologische Überzeugungen und Überzeugungen zum forschungsmethodischen Wissen zur Validierung herangezogen. Durch die Analyse von Zusammenhängen der benannten Variablen mit Leistungsmaßen für die Bereiche *Informationsauswahl* und *Bewertung von Studien* können weitere Argumente für die Konstruktvalidität erbracht werden.

2.4 Zielsetzung und Fragestellung der vorliegenden Studie

Ziel der vorliegenden Studie war es, Testinstrumente zur Messung der Leistung bei *Informationsauswahl* und *Bewertung von Studien* im Rahmen des evidenzbasierten Argumentierens zu entwickeln und zu validieren. Nach der Darstellung des Instrumentes und insbesondere der Operationalisierung der Teilbereiche *Informationsauswahl* und *Bewertung von Studien* werden Argumente für die Validität zusammengetragen (Kane, 1992). Die Fragestellung der vorliegenden Beitrags lautet:

Welche empirischen Argumente können für die Inhalts- und Konstruktvalidität des entwickelten Verfahrens zur Messung der *Informationsauswahl* und der *Bewertung von Studien* im Kontext von evidenzbasiertem Argumentieren angeführt werden?

3. Methode

3.1 Erhebungsinstrumente

Szenario des Kompetenztests

Für die Erfassung der Kompetenz im evidenzbasierten Argumentieren wurde ein computerbasiertes Szenario entwickelt. Computerbasierte Szenarien bieten im Gegensatz zu Papier-und-Bleistift-basierten Instrumenten den Vorteil, komplexe Interaktionsmuster messbar und quantifizierbar zu machen sowie die Messung von komplexen Kompetenzen in kontextualisierten und möglichst realitätsnahen Problemlösesituationen durchzuführen (siehe Diskussion bei Frey & Hartig, 2013, S. 54). Innerhalb des Szenarios sollen die Probanden (in der Rolle einer Lehrkraft) im Rahmen eines wirtschaftsfinanzierten „business start“-Wettbewerbs, in dem Schüler ein eigenes Geschäftsmodell entwickeln und präsentieren lernen sollen, ein Präsentationstraining für eine Gruppe mit 15-jährigen Schülerinnen und Schülern vorbereiten. Im Laufe der Trainingsplanung gilt es, zwei Entscheidungen zu treffen: (1) Die Einleitung und Eröffnung einer Präsentation soll mit Videobeispielen trainiert werden. Die Probanden sollen entscheiden, ob allein vorbildliche oder eine Kombination von vorbildlichen und fehlerhaften Videobeispielen eingesetzt werden sollten. (2) Im zweiten Teil des Trainings sollen Grundlagen aus der Rhetorik vermittelt werden. Hier ist von den Probanden zu entscheiden, ob in diesem Fall die Methode des „Gruppenpuzzles“ oder die Einzelarbeit eingesetzt werden sollte. Der fiktive Geldgeber für das Trainingsprogramm besteht auf Evidenzbasierung; daher ist es notwendig, nach geeigneten wissenschaftlichen Studien zu suchen und insbesondere deren Qualität zu bewerten.

Zu beiden Entscheidungen liegen ausreichend Forschungsergebnisse vor, sodass bei der Entwicklung der Materialien aus einer Menge von wissenschaftlichen Studien ausgewählt werden konnte. Zudem sind die Inhaltsbereiche für unterschiedliche pädagogische Settings relevant (z. B. für angehende Lehrpersonen an Schulen oder auch für angehende Weiterbildnerinnen und Weiterbildner). Die ausgewählten Studien wurden in Bezug auf Fragestellungen, Methodik und Ergebnisse nicht verändert. Es wurden strukturierte Kurzfassungen der Originalarbeiten erstellt, die sich in einer anderen Studie als leichter zugänglich und besser verständlich erwiesen haben (Hetmanek et al., 2014). Alle Materialien wurden in deutscher Sprache präsentiert.

Die aus den theoretischen Vorüberlegungen abgeleiteten Teilschritte für die evidenzbasierte Entscheidungsfindung, die in unserer Testumgebung entsprechend umgesetzt wurden, sind in Tabelle 1 (inkl. Bearbeitungszeiten) dargestellt.

Pilotierungen ergaben, dass nach der Bearbeitung von zwei Texten die Sicherheit und Vertrautheit mit dem Bewertungsraster deutlich zunimmt. Daher wurde den Probanden vor der Bearbeitung der eigentlichen Testaufgabe eine Übungsentscheidung vorgelegt, anhand derer die Bewertung von Studien an zwei wissenschaftlichen Texten erprobt werden konnte.

Im Rahmen dieses Beitrags fokussieren wir die Instrumente für die Erfassung der Teilkompetenzen 1b, 1c und 2 (siehe Tab. 1).

	Teilkompetenz	vorgesehene Zeit
1a	Informationssuche	2 × 3 min
1b	Informationsauswahl anhand einer Liste mit bibliografischen Angaben	2 × 4 min
1c	Informationsauswahl anhand einer Liste mit Abstracts	2 × 7 min
2	Bewertung von Studien anhand von strukturierten Kurzfassungen	2 × 20 min (4 × 5 min)
3	evidenzbasierte Entscheidung und Argumentation	2 × 6 min

Tab. 1: Übersicht über Testteile und dafür vorgesehene Testzeiten

Informationsauswahl

Der Testteil zur *Informationsauswahl* umfasst bei jeder der beiden Entscheidungen zwei Teilaufgaben. Bei der *ersten Teilaufgabe* werden den Probanden bei jeder Entscheidung die zentralen bibliografischen Angaben (Autoren, Titel, Zeitschrift/Medium, Publikationsjahr) zu jeweils zehn Publikationen in einer Liste präsentiert, die in ihrem Aufbau den Trefferlisten sowohl üblicher digitaler Fachbibliografien (bspw. ERIC) als auch Suchmaschinen im Internet (z. B. Google Scholar) ähnelt (siehe Abb. 1). In den Listen sind jeweils vier hoch relevante, zwei eher relevante, zwei weniger relevante und zwei nicht relevante Publikationen enthalten. Die Probanden werden aufgefordert, die Studien auf einer Skala von 1 (auf keinen Fall) bis 4 (auf jeden Fall) im Hinblick darauf zu beurteilen, ob sie weitere Informationen darüber erhalten möchten. Die Ratings der Probanden werden mit einer im Projekt entwickelten Musterlösung abgeglichen. Wertet ein Proband eine hoch relevante Studie ebenfalls als hoch relevant (Antwort: auf jeden Fall), erhält er drei Punkte; weicht er mit seiner Einschätzung von der Musterlösung ab, erhält er für jeden Schritt Abstand einen Punkt weniger. Der maximal mögliche Wertebereich für ein Item reicht demnach von null bis drei Punkten. Der Gesamtwert für die erste Teilaufgabe der Informationsauswahl wird gebildet, indem die Bewertungen über die jeweils zehn Items bei den beiden Entscheidungen addiert werden. Auf dieser Skala können somit Werte bis 60 erzielt werden, wobei höhere Werte eine höhere Kompetenz in Bezug auf den ersten Teil der Informationsauswahl anzeigen.

Bei der *zweiten Teilaufgabe* werden den Probanden unabhängig von ihrer Auswahl beim vorhergehenden Schritt die Originalabstracts von sechs der zehn Publikationen, die bereits in der zuvor gezeigten Liste enthalten waren, vorgelegt (siehe Abb. 2). Vier dieser Publikationen sind für die vorliegende Entscheidung jeweils hoch relevant, zwei davon dagegen weniger relevant. Die Probanden werden daraufhin gebeten, jede der sechs Publikationen auf der Grundlage dieses erweiterten Informationsstandes auf einer Skala von 1 (stimme gar nicht zu) bis 7 (stimme völlig zu) im Hinblick darauf zu beurteilen, ob sie für die anstehende Entscheidung eine wichtige Rolle spielt. Die Ratings der Probanden werden einzeln mit der Musterlösung verglichen. Beurteilt ein Proband eine nicht relevante Studie ebenfalls als nicht relevant (Wert zwischen 1 und 3) oder eine relevante Studie als relevant (Wert zwischen 5 und 7), erhält er zwei Punkte; ent-

Einzelarbeit oder Gruppenpuzzle

Nach einigen Suchanfragen erhalten Sie die untenstehende Trefferliste. Bitte gehen Sie die Treffer durch und überlegen Sie, zu welcher der Studien Sie detailliertere Informationen haben möchten.

Wie schätzen Sie die Treffer ein?

Bitte wählen Sie eine der Antwortmöglichkeiten aus:

Auf Grundlage dieser Informationen möchte ich ...

... auf keinen Fall

... eher nicht

... eher schon

... auf jeden Fall

... mehr über die Studie wissen und eine Zusammenfassung ansehen.

Fördern Schulnoten die Motivation? Eine quasi-experimentelle Studie zum Einfluss der Benotungserwartung auf selbst berichtete und verhaltensnah erhobene Motivationsqualitäten

Hänze, M., Berger, R. & Bianchy, K - Psychologie in Erziehung und Unterricht - 2009

☐ auf keinen Fall

☐ eher nicht

☐ eher schon

☐ auf jeden Fall

Prozesse und Effekte „Kooperativen Lernens“ im Sportunterricht

Bähr, I., Prohl, R. & Gröben, B. - Unterrichtswissenschaft - 2008

☐ auf keinen Fall

☐ eher nicht

☐ eher schon

☐ auf jeden Fall

Der Experteneffekt: Grenzen kooperativen Lernens in der Primarstufe?

Borsch, F., Gold, A., Kronenberger, J. & Souvignier, E. - Unterrichtswissenschaft - 2007

☐ auf keinen Fall

☐ eher nicht

☐ eher schon

☐ auf jeden Fall

Abb. 1: Screenshot der Teilaufgabe Informationsauswahl anhand einer Liste mit bibliografischen Angaben

scheidet er sich für die Mitte, erhält er einen Punkt. Schätzt der Proband eine Studie entgegen der Musterlösung ein, erhält er null Punkte. Der Gesamtwert für die zweite Teilaufgabe der Informationsauswahl wird gebildet, indem die Bewertungen über die jeweils sechs Items bei beiden Entscheidungen addiert werden. Die Skala umfasst somit Werte von 0 bis 24.

Beide Teilaufgaben der Informationsauswahl werden schließlich durch Addition zu einem Gesamtwert zusammengefasst. Insgesamt können für die Informationsauswahl somit maximal 84 Punkte erreicht werden, wobei höhere Werte eine höhere Kompetenz in Bezug auf die Informationsauswahl anzeigen.

Bewertung von Studien

In diesem Testteil werden den Probanden – wiederum unabhängig von ihrer vorherigen Auswahl – zu den vier relevanten unter den bereits bekannten Publikationen *strukturierte Kurzfassungen* vorgelegt (siehe Abb. 3). Dabei handelt es sich um ca. zweiseitige strukturierte Kurzfassungen wissenschaftlicher Originalveröffentlichungen mit standardisiertem Aufbau, der eng an die übliche Gliederung eines Berichts über eine empirische Untersuchung angelehnt ist. Jede dieser Studien wird von den Probanden im Hinblick

Einzelarbeit oder Gruppenpuzzle

Auswahl von Untersuchungen

Bitte sehen Sie sich nun zu den folgenden Untersuchungen die Zusammenfassungen an und geben Sie unten Ihre Einschätzung zu der Frage ab, ob Sie auf der Grundlage dieser Informationen davon ausgehen, dass die Untersuchung für die anstehende Entscheidung eine wichtige Rolle spielt:

Fördern Schulnoten die Motivation? Eine quasi-experimentelle Studie zum Einfluss der Benotungserwartung auf selbst berichtete und verhaltensnah erhobene Motivationsqualitäten

Hänze, M., Berger, R. & Bianchy, K - Psychologie in Erziehung und Unterricht - 2009

Zusammenfassung: Es wird der Frage nachgegangen, wie sich Schulnoten auf intrinsische und extrinsische Motivationsformen im kooperativen Fachunterricht Physik der 12. Klasse auswirken. Es nahmen 15 Physikkurse mit 293 Schülerinnen und Schülern an der Untersuchung teil. Die Kurse wurden zufällig der Bedingung mit oder ohne Benotungserwartung zugewiesen. Bei der Bedingung mit Benotungserwartung (N=130) ging die Leistung in einem Abschlusstest in die mündliche Note für das Schulhalbjahr ein; bei der Bedingung ohne Benotungserwartung (N=163) hatte der Abschlusstest keine weitere persönliche Konsequenz für die Schüler. Es wurde die Lernleistung, die selbstberichtete Motivation während des Unterrichts und die häusliche Nutzung einer Online-Lernplattform als verhaltensnahes Maß für Motivation erhoben. Während die Lernleistung mit großem Effekt durch die Benotungserwartung beeinflusst wurde, gab es keinen Effekt auf die selbstberichtete Motivation. Die Lernplattform wurde zur Testvorbereitung von der Gruppe mit Benotungserwartung deutlich stärker benutzt; der Effekt zeigte sich nach dem Test bei einer (nicht notenrelevanten) verpflichtenden Hausaufgabe nur noch tendenziell und verschwand bei einer freiwilligen häuslichen Zusatzaufgabe. Eine Pfadanalyse zeigte, dass der Effekt der Benotungserwartung auf die Leistung sowohl über das intensivere individuelle Lernen mit der Lernplattform als auch über eine bessere selbstberichtete Kommunikationsqualität in den Lerngruppen vermittelt wurde.

Auf Grundlage dieser Informationen gehe ich davon aus, dass die Untersuchung eine wichtige Rolle für die Entscheidung spielt.

Ich stimme gar
nicht zu



Ich stimme
völlig zu



Abb. 2: Screenshot der Teilaufgabe Informationsauswahl anhand einer Liste mit Abstracts

auf die theoretisch abgeleiteten zehn Qualitätsaspekte beurteilt. Jeder dieser zehn Qualitätsaspekte wird auf einer Rating-Skala von 1 (stimme überhaupt nicht zu) bis 9 (stimme völlig zu) bewertet. Vor der Bearbeitung der beiden eigentlichen Entscheidungsaufgaben und nach Einführung des Szenarios zu Beginn der Testsitzung werden die Probanden anhand strukturierter Kurzfassungen von zwei weiteren Studien ohne Bezug zu den beiden Entscheidungen mit dieser Aufgabenstellung vertraut gemacht.

Das erste und das letzte Ratingitem (Bedeutung der Studie für die Entscheidung insgesamt ohne bzw. mit Berücksichtigung der einzelnen Qualitätsaspekte) werden bei der Bildung einer Skala für die *Bewertung von Studien* nicht berücksichtigt. Für jeden Probanden wird für jede der insgesamt acht Studien aus beiden Entscheidungen die Produkt-Moment-Korrelation zwischen den acht Ratings des Probanden und den acht Ratings aus einer im Projekt entwickelten Musterlösung berechnet. Der Wert pro Studie kann also ein Minimum von -1 und ein Maximum von $+1$ annehmen. Zur Bildung der Gesamtskala für die *Bewertung von Studien* wird der Mittelwert dieser acht Korrelationskoeffizienten gebildet, wobei höhere Werte eine höhere Kompetenz in Bezug auf die *Bewertung von Studien* anzeigen.

Einzelarbeit oder Gruppenpuzzle

Beurteilung von Untersuchungen

Bitte lesen Sie sich die im Folgenden angezeigte Kurzfassung einer wissenschaftlichen Untersuchung aufmerksam durch und geben Sie dann an, inwiefern sie den unten stehenden Aussagen zustimmen.

Dafür sind insgesamt etwa fünf Minuten Bearbeitungszeit vorgesehen.

Das Gruppenpuzzle im Physikunterricht der Sekundarstufe II – Einfluss auf Motivation, Lernen und Leistung

M. Hänze und R. Berger

In dieser Studie wurde untersucht, ob die grundlegenden Bedürfnisse nach sozialer Eingebundenheit, Kompetenz- und Autonomieerleben beim Lernen nach der Methode des Gruppenpuzzles in höherem Ausmaß befriedigt werden als im Frontalunterricht, und ob infolgedessen ein höherer Lernerfolg erzielt wird.

Methodik

Teilnehmer. In die Studie wurden Daten von 61 Zwölfklässlern aus acht Physik-Grundkursen an fünf Schulen einbezogen, die bei zwei Unterrichtseinheiten zu den Themen „Rasterelektronenmikroskop“ und „Mikrowellenofen“ in zwei verschiedenen Halbjahren anwesend waren.

Untersuchungsplan. Bei jeder Unterrichtseinheit wurde ein Teil der Grundkurse nach der Methode des Gruppenpuzzles unterrichtet, im anderen Teil fand Frontalunterricht statt. Bei der zweiten Unterrichtseinheit wurde die Zuordnung der Kurse zu den Bedingungen umgekehrt.

Unterrichtseinheiten und Ablauf. Die Unterrichtseinheit zum Thema „Rasterelektronenmikroskop“ umfasste vier Schulstunden, die zum Thema „Mikrowellenofen“ drei Schulstunden. Vor jeder Unterrichtseinheit wurde das Vorwissen zum Thema erfasst. Im Anschluss wurde jeweils ein unbenoteter und unangekündigter Abschlusstest durchgeführt.

	Ich stimme überhaupt nicht zu								Ich stimme völlig zu	
Die Studie spielt für die aktuelle Entscheidung eine wichtige Rolle.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aus dem Text geht hervor, dass die für die Entscheidung relevante didaktische Gestaltungsvariante untersucht wurde.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aus dem Text geht hervor, dass die für die Entscheidung relevante didaktische Gestaltungsvariante im Kontext der aktuellen Entscheidung leicht umsetzbar ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aus dem Text geht hervor, dass die untersuchten Erfolgskriterien mit den Erfolgskriterien bei der aktuellen Entscheidung überein stimmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aus dem Text geht hervor, dass die Erfolgskriterien mit geeigneten Verfahren gemessen wurden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aus dem Text geht hervor, dass die Rahmenbedingungen mit den Rahmenbedingungen bei der aktuellen Entscheidung in den wichtigen Merkmalen übereinstimmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aus dem Text geht hervor, dass die Teilnehmer mit der Zielgruppe bei der aktuellen Entscheidung in den wichtigen Merkmalen übereinstimmen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aus dem Text geht hervor, dass die für die Entscheidung relevanten Ergebnisse statistisch belastbar sind.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Soweit ersichtlich, sind die für die Entscheidung relevanten Effekte eindeutig auf die untersuchte Gestaltungsvariante zurückzuführen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Unter Berücksichtigung der zuvor bewerteten Aspekte spielt die Studie für die aktuelle Entscheidung eine wichtige Rolle.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Abb. 3: Screenshot der Teilaufgabe Bewertung von Studien

Hintergrundvariablen

Allgemeine kognitive Fähigkeiten. Diese wurden zum einen mit einer selbst zusammengestellten Kurzfassung der verbalen Analogien aus dem *Intelligenz-Struktur-Test* (I-S-T 2000R; Liepmann, Beauducal, Brocke & Amthauer, 2007) sowie einer validierten Kurzfassung des *Raven's Advanced Progressive Matrices Test* (Bors & Stokes, 1998; Raven & Court, 1988) erfasst.

Wissen. Es wurden zwei Wissensbereiche berücksichtigt: Der Fragebogen zum *pädagogischen Wissen* wurde im Projekt entwickelt und umfasst Fragen zu Theorien des Lehrens und Lernens mit Aufgaben im Single-Choice-Format. Ein Beispielitem lautet: „Welches der folgenden Unterrichtsmerkmale gehört nicht zu den Qualitätsmerkmalen guten Unterrichts nach Helmke?“ (richtige Antwort: Präzision; Distraktoren: Klarheit und Strukturiertheit, Aktivierung, Konsolidierung und Sicherung). Die Skala *forschungsmethodisches Wissen* wurde ebenfalls selbst entwickelt und erfasst das Wissen der Probanden in Hinblick auf methodische Merkmale wissenschaftlicher Studien mit Aufgaben im Single-Choice-Format. Ein Beispielitem lautet: „Was versteht man unter „Reliabilität“? (richtige Antwort: die Messgenauigkeit eines Messinstruments; Distraktoren: die Aussagekraft eines Tests für das zu erfassende Merkmal, das Ausmaß, in dem Testergebnisse von den durchführenden Personen unabhängig sind, die Belastbarkeit der Ergebnisse einer Untersuchung).

Wissenschaftliches Denken (scientific reasoning). Die Skala erfasst die Fähigkeit der Probanden, wissenschaftliche Schlussfolgerungen zu ziehen. Für die Erfassung dieser Fähigkeit wurden Items im Single-Choice-Format aus dem Instrument von Lippman (2012) verwendet.

Kritisches Denken (critical thinking). Die Skala umfasst die Fähigkeit des induktiven Schlussfolgerns und Reflektierens von Evidenz. Die Aufgaben im Single-Choice-Format wurden ebenfalls von Lippman (2012) übernommen. Die Skalen zum wissenschaftlichen und kritischen Denken wurden jeweils durch die Berechnung der Summe der Punkte bei den Einzelitems gebildet.

Epistemologische Überzeugungen. Diese wurden mit einer deutschen Fassung des *Discipline-Focused Epistemological Beliefs Questionnaire* (Hofer, 2000) auf einer Skala von 1 (stimme nicht zu) bis 5 (stimme zu) erfasst. Zur Berechnung des Gesamtscores wurde ein Mittelwert gebildet.

Überzeugungen zu forschungsmethodischem Wissen. Diese Skala erfasst Selbsteinschätzungen zum Verständnis forschungsbezogener Fachbegriffe (adaptiert nach Bauer & Prenzel, 2012; Jette et al., 2003) auf einer Skala von 1 (ich kenne den Begriff nicht) bis 4 (ich verstehe den Begriff und könnte ihn anderen erklären). Ein Beispielitem lautet: „Wenn Sie einen bildungswissenschaftlichen Forschungsbericht lesen, könnten darin unter anderem folgende Fachbegriffe auftauchen: a. Evidenz, b. Randomisiertes Kontrollgruppendesign, c. Quasi-Experiment, d. Meta-Analyse, e. Repräsentative Stichprobe, f. Effektstärke. Wie gut schätzen Sie Ihr Verständnis dieser Fachbegriffe ein?“. Die statistischen Kennwerte der Variablen im Vortest sind in Tabelle 2 aufgelistet.

Variablen	Anzahl Items	M	SD	Cronbachs α	Min	Max
Allgemeine kognitive Fähigkeiten – verbale Analogien	10	5.86	1.90	.53	0	10
Allgemeine kognitive Fähigkeiten – figurale Analogien	12	6.69	2.54	.71	0	12
Forschungsmethodisches Wissen	16	3.53	1.65	.46	0	8
Pädagogisches Wissen	10	5.07	1.72	.31	0	10
Wissenschaftliches Denken	10	4.95	2.43	.65	0	10
Kritisches Denken	15	8.11	2.34	.35	0	13
Epistemologische Überzeugungen	18	2.32	.39	.69	–	–
Überzeugungen zu forschungs-methodischem Wissen	10	2.96	.58	.78	–	–

Tab. 2: Skalenskennwerte der Hintergrundvariablen

3.2 Ablauf der Untersuchung

Vor der Durchführung des Kompetenztests wurden alle Probanden gebeten, einen Online-Test für die Hintergrundvariablen (siehe Tab. 2) sowie demografische Angaben auszufüllen. Der entsprechende Link zum Online-Test wurde per Mail an die Probanden versendet. Die Bearbeitung des Kompetenztests wurde mit allen Studierenden in Präsenz mit je einer Versuchsleiterin bzw. einem Versuchsleiter unter Verwendung eines standardisierten Leitfadens durchgeführt. Die mittlere Bearbeitungszeit betrug 74 Minuten für den Haupttest und 57 Minuten für den Vortest.

3.3 Stichprobe

Die folgenden Analysen beruhen auf einer Stichprobe von insgesamt 341 Studierenden. Von diesen waren 78.8% weiblich und 17.4% männlich (3.8% fehlende Angaben). Das durchschnittliche Alter lag bei $M = 23.13$ Jahren ($SD = 5.27$). Unter den Teilnehmern befanden sich 157 Masterstudierende in den Bildungswissenschaften, 109 Studierende eines Bachelorstudiengangs Pädagogik/Bildungswissenschaft, 36 Lehramtsstudierende und 39 Promotionsstudierende aus den Bereichen Bildungswissenschaften, Pädagogik und Psychologie. Die Master-, Bachelor- und Lehramtsstudierenden stammten von der Bergischen Universität Wuppertal und der Ludwig-Maximilians-Universität München. Die befragten Promotionsstudierenden wurden deutschlandweit rekrutiert und absolvierten die Testsitzung online. Damit ist die vergleichsweise geringe Stichprobengröße in dieser Gruppe zu erklären. Der Großteil der Teilnehmer (86.9%) hat Deutsch als Muttersprache.

4. Ergebnisse

4.1 Inhaltsvalidität

Informationsauswahl

Die inhaltliche Validierung der Teilkompetenz beruht auf der Befragung von Experten. Hierfür wurden die konstruierten Trefferlisten in zufälliger Reihenfolge zusammen mit den nötigen Hintergrundinformationen aus dem Szenario elf wissenschaftlichen Mitarbeitern (Doktoranden, Postdoktoranden, Professoren) am Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie der Universität München zur Bearbeitung entsprechend der Testaufgabe vorgelegt. Als Kennwert für die Übereinstimmung der Beurteilungen wurde eine Intraklassenkorrelation über alle Beurteiler und über beide Entscheidungen hinweg ermittelt ($ICC_{just.MW} = .95$).

Bewertung von Studien

Für die inhaltliche Validierung der Teilkompetenz *Bewertung von Studien* wurden Experten in verschiedener Weise eingebunden: Zunächst wurde die Teilaufgabe von zwei wissenschaftlichen Mitarbeitern der LMU München mit einschlägigen Kenntnissen im Bereich der Evidenzbasierung bearbeitet. Die initiale Beurteilerübereinstimmung wurde anhand einer Intraklassenkorrelation berechnet ($ICC_{just.MW} = .78$). Die Lösungen der Experten wurden verglichen, und differierende Bewertungen wurden konsensuell ausdiskutiert. Diese Musterlösung wurde mit einem weiteren Projektmitarbeiter und Experten diskutiert und nach Diskussion der Hauptpunkte noch einmal angeglichen.

4.2 Konstruktvalidität

In Bezug auf die Konstruktvalidierung wurde die interne Konsistenz (Cronbachs α) der beiden Teilkompetenzen *Informationsauswahl* und *Bewertung von Studien* sowie der Zusammenhang dieser beiden Teilskalen ermittelt. Die interne Konsistenz der Teilkompetenz der *Informationsauswahl* lag bei Cronbachs $\alpha = .65$, die der Teilkompetenz *Bewertung von Studien* bei Cronbachs $\alpha = .53$. Die beiden Teilkompetenzen korrelierten signifikant miteinander ($r = .22, p < .01$). Zudem wurden Korrelationen zwischen den einbezogenen Teilkompetenzen *Informationsauswahl* und *Bewertung von Studien* und den erhobenen Hintergrundvariablen berechnet. Die Ergebnisse (siehe Tab. 3) zeigen signifikante Korrelationen der Teilkompetenz *Informationsauswahl* mit den verbalen Analogien aus dem Intelligenz-Struktur-Test, der Fähigkeit des kritischen Denkens, den Überzeugungen zu forschungsmethodischem Wissen, dem pädagogischen Wissen und der Fähigkeit des wissenschaftlichen Denkens. Die Teilkompetenz *Bewertung von Studien* weist einen signifikanten Zusammenhang mit allen einbezogenen Variablen auf. Insbesondere die verbalen Analogien aus dem Intelligenz-Struktur-Test, die epistemologischen Überzeugungen und das wissenschaftliche Denken weisen die höchsten positiven Zusammenhänge mit dieser Teilkompetenz auf.

	Informationsauswahl	Bewertung von Studien
Allgemeine kognitive Fähigkeiten		
Verbale Analogien	.21**	.25**
Figurale Analogien	.12	.15*
Forschungsmethodisches Wissen		
Pädagogisches Wissen	.13*	.19**
Wissenschaftliches Denken		
Kritisches Denken	.15*	.23**
Überzeugungen zu forschungs- methodischem Wissen		
Epistemologische Überzeugungen	.08	.24**

** $p < .01$; * $p < .05$

Tab. 3: Korrelation der Teilkompetenzen Informationsauswahl und Bewertung von Studien mit den Hintergrundvariablen

Um gemeinsame Varianzanteile der Hintergrundvariablen angemessen zu berücksichtigen, wurden anschließend multiple Regressionsanalysen mit den beiden Teilkompetenzen als abhängigen Variablen und den übrigen Variablen als Prädiktoren durchgeführt (siehe Tab. 4). Die Hintergrundvariablen wurden dafür in vier theoretisch gebildeten Blöcken (hierarchisch) eingeschlossen, um den individuellen Varianzanteil der einzelnen Blöcke am Gesamtmodell zu erhalten: Der erste Block bildet die allgemeinen kognitiven Fähigkeiten ab, der zweite Block die Wissensvariablen und der dritte Block die Fähigkeiten zum wissenschaftlichen und kritischen Denken. Im vierten Block wurden die übrigen Variablen Überzeugungen zu forschungsmethodischem Wissen sowie epistemologische Überzeugungen zusammengefasst. Für die Teilkompetenz *Informationsauswahl* zeigt sich, dass allgemeine kognitive Fähigkeiten fünf Prozent der Varianz aufklären. Weitere Prädiktoren spielen für die Vorhersage der Informationsauswahl keine statistisch belastbare Rolle. Eine Regressionsanalyse mit allen Prädiktoren klärt einen Varianzanteil von acht Prozent auf ($F(8, 232) = 2.69, p < .01$).

Bei der Teilkompetenz *Bewertung von Studien* zeigt sich ein anderes Bild: Allgemeine kognitive Fähigkeiten klären einen signifikanten Varianzanteil von sieben Prozent auf. Forschungsmethodisches und pädagogisches Wissen klären gemeinsam weitere zwei Prozent der Varianz auf; dieser Zuwachs an aufgeklärtem Varianzanteil ist jedoch nicht signifikant. Die Fähigkeiten des wissenschaftlichen und des kritischen Denkens erklären gemeinsam signifikant weitere zwei Prozent der Varianz der Teilkompetenz *Bewertung von Studien*. Epistemologische Überzeugungen leisten einen signi-

Prädiktoren	Informationsauswahl				Bewertung von Studien			
	Modell 1	Modell 2	Modell 3	Modell 4	Modell 1	Modell 2	Modell 3	Modell 4
Allgemeine kognitive Fähigkeiten:								
Verbale Analogien	.19**	.19*	.18*	.17*	.26**	.19**	.16*	.17*
Figurale Analogien	.34	.02	-.02	-.01	.06	.04	-.02	-.04
Forschungsmethodisches Wissen		-.03	-.01	-.04		.09	.09	.09
Pädagogisches Wissen		.08	.07	.04		.10	.08	.03
Wissenschaftliches Denken			.04	.02			.12	.11
Kritisches Denken			.13(*)	.11			.10	.09
Überzeugungen zu forschungsmethodischem Wissen				.13				.04
Epistemologische Überzeugungen				.01				.17**
Änderung in R ²	.05**	.01	.02	.01	.07**	.02	.02*	.03*

** $p < .01$; * $p < .05$; (*) $p < .07$

Tab. 4: Standardisierte Regressionskoeffizienten aus multiplen hierarchischen Regressionen mit den Teilkompetenzen als abhängige und den Hintergrundvariablen als unabhängige Variablen

fikanten Beitrag zur Vorhersage der Teilkompetenz *Bewertung von Studien* und klären gemeinsam mit den Überzeugungen zu forschungsmethodischem Wissen weitere drei Prozent der Varianz auf. Insgesamt können durch die Regressionsgleichung mit allen übrigen Variablen als Prädiktoren 14 Prozent der Varianz der Teilkompetenz *Bewertung von Studien* erklärt werden ($F(8, 231) = 4.54, p < .01$).

5. Diskussion

Die Umsetzung einer stärker evidenzbasierten Praxis im Bildungsbereich auf der Ebene von individuellen Entscheidungen stellt hohe kognitive Anforderungen an pädagogische Praktiker (Bromme et al., 2014, S. 8). Ein ganzes Bündel von Fertigkeiten muss hierfür eingesetzt werden. Zentral sind die Bereiche der *Informationsauswahl* und der *Bewertung von Studien*. Für diese Teilkompetenzen der Kompetenz im evidenzbasierten Argumentieren wurde im vorliegenden Beitrag anhand von kognitionspsychologischen Erkenntnissen ein theoretisches Modell entwickelt, und anschließend wurden empirische Argumente für die Validität der entwickelten Instrumente gesammelt. Die zentralen Ergebnisse dieser Analysen lauten:

- 1) Die Teilkompetenzen *Informationsauswahl* und *Bewertung von Studien* sind empirisch voneinander unterscheidbar. Der schwache, wenngleich signifikante Zusammenhang zwischen den beiden Teilkompetenzen belegt empirisch das theoretisch erwartete Befundmuster der voneinander trennbaren Teilkompetenzen *Informationsauswahl* und *Bewertung von Studien*. Diese beruhen auf hinreichend unterschiedlichen kognitiven Aktivitäten. Zudem weisen sie einen erwartbaren Zusammenhang mit Maßen für allgemeine kognitive Fähigkeiten auf, sind aber dennoch deutlich von diesen unterscheidbar. Dieses Ergebnis kann als Indikator für die Konstruktvalidität gewertet werden (Cronbach & Meehl, 1955).
- 2) Betrachtet man die Korrelationen zwischen den beiden untersuchten Teilkompetenzen und den Hintergrundvariablen, bietet sich folgendes Bild: Beide Teilkompetenzen hängen mit allgemeinen (insbesondere verbalen) kognitiven Fähigkeiten zusammen. Bei der *Informationsauswahl* scheint ein Zusammenhang mit dem pädagogischen Wissen und bei der *Bewertung von Studien* zusätzlich ein Zusammenhang mit dem forschungsmethodischen Wissen zu bestehen. Diese Befunde wären in Anbetracht der unterschiedlichen Anforderungen bei den beiden Teilkompetenzen plausibel und theoretisch erwartbar (Brand-Gruwel et al., 2005): Für die *Informationsauswahl* dürfte pädagogisches Wissen wichtig sein – z. B. für das Verständnis von pädagogischen Fachbegriffen. Bei der *Bewertung von Studien* könnte darüber hinaus forschungsmethodischem Wissen eine wichtige Rolle zukommen: Je höher das forschungsmethodische Wissen ausgeprägt ist, desto besser dürfte eine Bewertung von Studien einschließlich ihrer forschungsmethodischen Qualität vorgenommen werden können. Diese Zusammenhänge mit dem pädagogischen und dem forschungsmethodischen Wissen verschwinden jedoch allesamt, wenn die allgemeinen kognitiven Fähigkeiten statistisch kontrolliert werden. Möglicherweise leistet Wissen bei der Anwendung der beiden Teilkompetenzen keinen Beitrag über das Ausmaß hinaus, in dem es selbst Niederschlag allgemeiner kognitiver Fähigkeiten ist. Derartige Erklärungsansätze erfordern jedoch weitere empirische Untersuchungen. Ähnliches gilt für die Rolle des wissenschaftlichen und des kritischen Denkens.

Zusammenfassend wollen wir folgende Punkte noch einmal festhalten: In diesem Beitrag konnten empirische Hinweise dafür gesammelt werden, dass die Kompetenz im evidenzbasierten Argumentieren aus wenigstens zwei zu differenzierenden Teilkompetenzen besteht. Die entwickelten Testinstrumente stellen eine Weiterentwicklung des zugrunde liegenden Modells des *Information Problem Solving* dar (Lazonder & Rouet, 2008), in dem die Teilkompetenz der *Informationsauswahl* um die Teilkompetenz zur systematischen *Bewertung von Studien* ergänzt wurde. Das Bewältigen der komplexen Anforderungen bei der *Informationsauswahl* und der *Bewertung von Studien* kann nicht durch anforderungsspezifisches Wissen, allgemeine kognitive Fähigkeiten und epistemologische Überzeugungen allein erklärt werden. Es ist plausibel anzunehmen, dass – das zeigen die Erfahrungen im Bereich der Medizin – die komplexe Kompetenz, evidenzbasiert zu argumentieren, gefördert werden kann. Dies auch für den Bildungsbereich nachzuweisen, sollte Gegenstand zukünftiger Forschung sein. Mit Blick auf die

Standards für Lehrerbildung der KMK hat die vorliegende Studie Hinweise geliefert, dass die untersuchten Teilkompetenzen differenziert betrachtet werden sollten (KMK, 2014). Außerdem legen die gewonnenen Erkenntnisse nahe, dass noch weitere Teilkompetenzen der Kompetenz im evidenzbasierten Argumentieren theoretisch ausdifferenziert und empirisch bestätigt werden sollten (siehe Tab. 1, z. B. die Rechtfertigung einer evidenzbasierten Entscheidung).

Nächste Schritte zur Weiterentwicklung der evidenzbasierten Praxis im Bildungsbereich könnten durch die Abstimmung von Studieninhalten und die Integration entsprechender Inhalte in Maßnahmen zur Qualitätssicherung und Weiterbildung von pädagogischen Fachkräften angegangen werden. Weitere Stärkung könnte die Idee evidenzbasierter pädagogischer Praxis erfahren, wenn es zukünftiger Forschung gelänge, die ganze Wirkungskette von der Planung bis zur Umsetzung in der Lehre und den Resultaten dieser Praxis bei den Lernenden in den Blick zu nehmen.

Literatur

- Amelang, M. (2000). Intelligenz. In M. Amelang (Hrsg.), *Enzyklopädie der Psychologie. Differentielle Psychologie und Persönlichkeitsforschung. Bd. 4: Determinanten individueller Unterschiede* (S. 245–328). Göttingen: Hogrefe.
- Astleitner, H., Brünken, R., & Zander, S. (2002). Können Schüler und Lehrer kritisch denken? Lösungserfolg und -strategien bei typischen Aufgaben. *Salzburger Beiträge zur Erziehungswissenschaft*, 6(2), 51–61.
- Bauer, J., & Prenzel, M. (2012). *Scientific and professional journals as resource for teachers' professional learning and evidence-based practice*. Vortrag auf der EARLI SIG 14 (Learning and Professional Development) Conference „Learning in transition“, Antwerpen, 22. bis 24. August 2012.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K., & Weiß, M. (Hrsg.) (2002). *PISA 2000: Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Beelmann, A. (2014). Möglichkeiten und Grenzen systematischer Evidenzkumulation durch Forschungssynthesen in der Bildungsforschung. *Zeitschrift für Erziehungswissenschaft*, 17(4), 55–78.
- Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, 58(3), 382–398.
- Brand-Gruwel, S., Wopereis, I. G. J. H., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, 21, 487–508.
- Brand-Gruwel, S., Wopereis, I. G. J. H., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education*, 53(4), 1207–1217.
- Bromme, R. (2014). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens* (2. Aufl.). Münster: Waxmann.
- Bromme, R., Prenzel, M., & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft*, 17(4), 3–54.

- Brown, N. J. S., Furtak, E. M., Timms, M., Nagashima, S. O., & Wilson, M. (2010). The Evidence-Based Reasoning Framework: Assessing scientific reasoning. *Educational Assessment*, 15(3/4), 123–141.
- Cook, B. G., Smith, G. J., & Tankersley, M. (2012). Evidence-based practices in education. In K. R. Harris, S. Graham & T. Urdan (Hrsg.), *APA Educational Psychology Handbook. Bd. 1: Theories, constructs, and critical issues* (S. 495–528). Washington, D. C.: American Psychological Association.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-Experimentation: Design and analysis issues for field settings*. Chicago: Rand-McNally.
- Cook, T. D., & Gorard, S. (2007). What counts and what should count as evidence. In OECD (Hrsg.), *Evidence in education: Linking research and policy* (S. 33–49). Paris: OECD.
- Cronbach, L. J. (1982). *Designing evaluations of educational and social programmes*. San Francisco: Jossey-Bass.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Evidence-Based Medicine Working Group (EBMWG) (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *Journal of the American Medical Association*, 268(17), 2420–2425.
- Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson (Hrsg.), *The Cambridge handbook of expertise and expert performance* (S. 41–67). New York: Cambridge University Press.
- Fischer, F., Waibel, M., & Wecker, C. (2005). Nutzenorientierte Grundlagenforschung im Bildungsbereich. *Zeitschrift für Erziehungswissenschaft*, 8(3), 427–442.
- Frey, A., & Hartig, J. (2013). Wann sollten computerbasierte Verfahren zur Messung von Kompetenzen anstelle von papier- und bleistift-basierten Verfahren eingesetzt werden? *Zeitschrift für Erziehungswissenschaft*, 16(1), 53–57.
- Goldman, S. R., Lawless, K. A., Pellegrino, J., Braasch, J., & Gomez, K. (2012). A technology for assessing multiple source comprehension: An essential skill of the 21st century. In M. C. Mayrath, J. Clarke-Midura & D. H. Robinson (Hrsg.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (S. 173–209). Charlotte: Information Age Publishing.
- Harden, M., Grant, J., Buckley, G., & Hart, I. R. (1999). BEME Guide No. 1: Best evidence medical education. *Medical Teacher*, 21(6), 553–62.
- Hartig, J., Frey, A., & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (2. Aufl., S. 143–172). Wiesbaden: Springer.
- Hetmanek, A., Wecker, C., Trempler, K., Kieseewetter, J., Hake, C., Gräsel, C., Fischer, M. R., & Fischer, F. (2014). *Lowering the bar for evidence-based practice in teaching – what structured abstracts may contribute*. Vortrag auf der EARLI SIG 11 (Teaching and Teacher Education) Conference „Practice-Oriented Teacher Learning and Professional Development“, Fraunchemsee, 16. bis 18. Juni 2014.
- Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology*, 25, 378–405.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140.
- Jenßen, L., Dunekacke, S., & Blömeke, S. (2015). Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. *Zeitschrift für Pädagogik*, 61. Beiheft, 11–31.
- Jette, D. U., Bacon, K., Batty, C., Carlson, M., Ferland, A., Hemingway, R. D., Hill, J., Ogilvie, L., & Volk, D. (2003). Evidence-based practice: Beliefs, attitudes, knowledge, and behaviors of physical therapists. *Journal of the American Physical Therapy*, 83, 786–805.

- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535.
- Klieme, E., Funke, J., Leutner, D., Reimann, P., & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz. Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47(2), 179–200.
- Ständige Konferenz der Kultusminister der Länder der Bundesrepublik Deutschland (2014). *Standards für die Lehrerbildung: Bildungswissenschaften*. http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf [05.08.2014].
- Lazonder, A. W., & Rouet, J.-F. (2008). Information problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior*, 24(3), 753–765.
- Liepmann, D., Beauducal, A., Brocke, B., & Amthauer, R. (2007). *I-S-T 2000 R* (2. Aufl.). Göttingen: Hogrefe.
- Lippman, J. P. (2012). *Improving and predicting novice reasoning about the evidentiary connection between psychological studies and theories* (unveröffentlichte Dissertation). Chicago: University of Illinois.
- Mayer, R. E. (2003). What causes individual differences in cognitive performance. In R. J. Sternber & E. L. Grigorenko (Hrsg.), *The psychology of abilities, competencies, and expertise* (S. 263–273). Cambridge: Cambridge University Press.
- Montori, V., & Guyatt, G. (2008). Progress in evidence-based medicine. *Journal of the American Medical Association*, 300(15), 1814–1816.
- Pant, H. A. (2014). Aufbereitung von Evidenz für bildungspolitische und pädagogische Entscheidungen: Metaanalysen in der Bildungsforschung. *Zeitschrift für Erziehungswissenschaft*, 17(4), 79–99.
- Raven, J. C., & Court, J. H. (1988). *Raven's Progressive Matrices und Vocabulary Scales. Manual. Teil 4: Advanced Progressive Matrices*. Set I & II. Frankfurt: Swets & Zeitlinger.
- Rosenberg, W. M. C., & Donald, A. (1995). Evidence based medicine: An approach to clinical problem-solving. *BMJ: British Medical Journal*, 310(6987), 1122–1126.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, S. W. (1996). Evidence based medicine: What it is and what it isn't. *BMJ: British Medical Journal*, 312 (7023), 71–72.
- Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence*, 4, 181–201.
- Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37(1), 5–14.
- Tugwell, P., Haynes, R. B., & Sackett, D. L. (1992). *Clinical epidemiology: A basic science for clinical medicine*. Boston: Little, Brown and Company.
- Wecker, C. (2013). How to support prescriptive statements by empirical research: Some missing parts. *Educational Psychology Review*, 25(1), 1–18.
- Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal*, 46(4), 1060–1106.
- Wopereis, I. G. J. H., & van Merriënboer, J. J. G. (2011). Evaluating text-based information on the World Wide Web. *Learning and Instruction*, 21(2), 232–237.
- Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, 20(1), 99–149.

Abstract: Following the model of medicine, evidence-based practice can be characterized as the conscientious use of the best currently available empirical evidence in professional practice. This is a complex challenge for educational practitioners. The goal of the present paper is to validate an instrument for the measurement of the required competence. The focus is on the two sub-competences of *information selection* and *evaluation of studies*. The participants were 341 students from degree programs in education. They completed a case-based online test for information selection and evaluation of studies. The results indicate that the two sub-competences can be separated and exhibit plausible correlations to other variables.

Keywords: Validation, Competency Assessment, Case-Based Decision, Evidence-Based Practice in Education, Empirical Studies

Anschrift der Autor(inn)en

Kati Trempler, Bergische Universität Wuppertal, School of Education,
Institut für Bildungsforschung, Gaußstraße 20, 42119 Wuppertal, Deutschland
E-Mail: trempler@uni-wuppertal.de

M. A. Andreas Hetmanek, Ludwig-Maximilians Universität München,
Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie,
Leopoldstraße 13, 80803 München, Deutschland
E-Mail: andreas.hetmanek@psy.lmu.de

PD Dr. Christof Wecker, Ludwig-Maximilians Universität München,
Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie,
Leopoldstraße 13, 80803 München, Deutschland
E-Mail: christof.wecker@psy.lmu.de

Dr. Jan Kiesewetter, Institut für Didaktik und Ausbildungsforschung in der Medizin,
Klinikum der Universität München, Ziemssenstraße 1, 80336 München, Deutschland
E-Mail: jan.kiesewetter@med.uni-muenchen.de

Mia Wermelt, Institut für Didaktik und Ausbildungsforschung in der Medizin,
Klinikum der Universität München, Ziemssenstraße 1, 80336 München, Deutschland
E-Mail: mia.wermelt@med.uni-muenchen.de

Prof. Dr. Frank Fischer, Ludwig-Maximilians Universität München,
Lehrstuhl für Empirische Pädagogik und Pädagogische Psychologie,
Leopoldstraße 13, 80803 München, Deutschland
E-Mail: Frank.Fischer@psy.lmu.de

Prof. Dr. Martin Fischer, Institut für Didaktik und Ausbildungsforschung in der Medizin,
Klinikum der Universität München, Ziemssenstraße 1, 80336 München, Deutschland
E-Mail: martin.fischer@med.uni-muenchen.de

Prof. Dr. Cornelia Gräsel, Bergische Universität Wuppertal, School of Education,
Institut für Bildungsforschung, Gaußstraße 20, 42119 Wuppertal, Deutschland
E-Mail: graesel@uni-wuppertal.de